Course Project 1

Marco Lopez

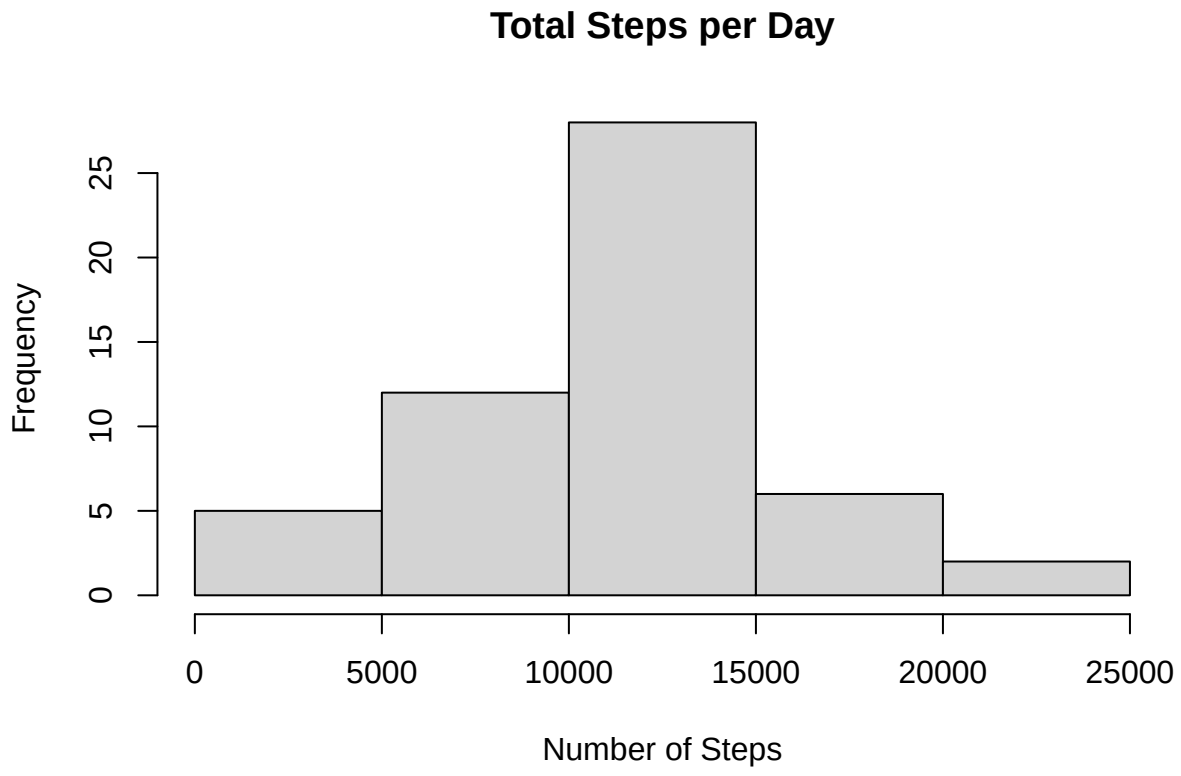
31/3/2022

Loading and preprocessing the data

```
Data <- read.csv(file="activity.csv", header=TRUE)
```

What is mean total number of steps taken per day?

Make a histogram of the total number of steps taken each day



Calculate and report the mean and median of the total number of steps taken per day

```
meanS <- mean(TStepforday$steps, na.rm = TRUE)
medD <- median(TStepforday$steps, na.rm = TRUE)
```

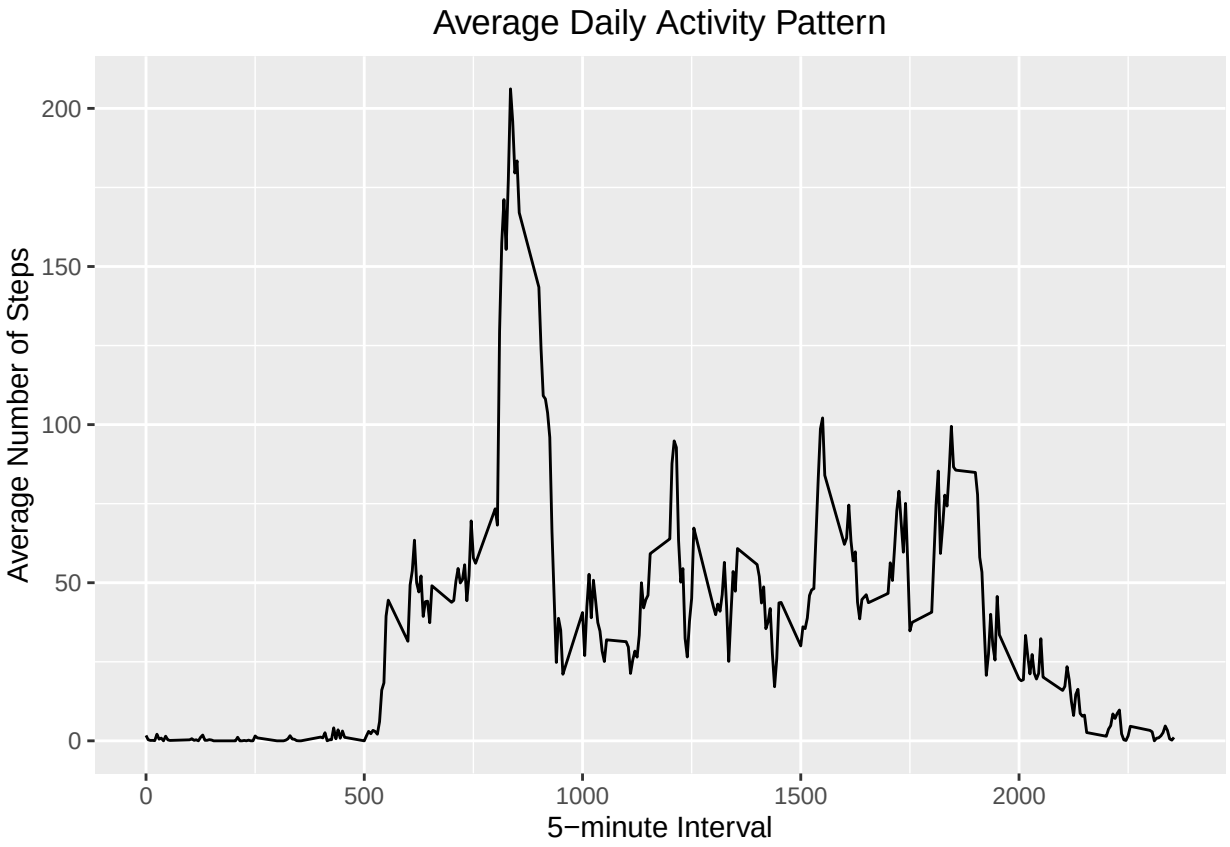
What is the average daily activity pattern?

Make a time series plot

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.2
```

```
meanSteps <- aggregate(steps ~ interval, Data, mean)
ggplot(data = meanSteps, aes(x = interval, y = steps)) + geom_line() + ggtitle("Average Daily Activity P
```



Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps?

```
maxstepinterval <- meanSteps[which.max(meanSteps$steps),]
```

Calculate and report the total number of missing values

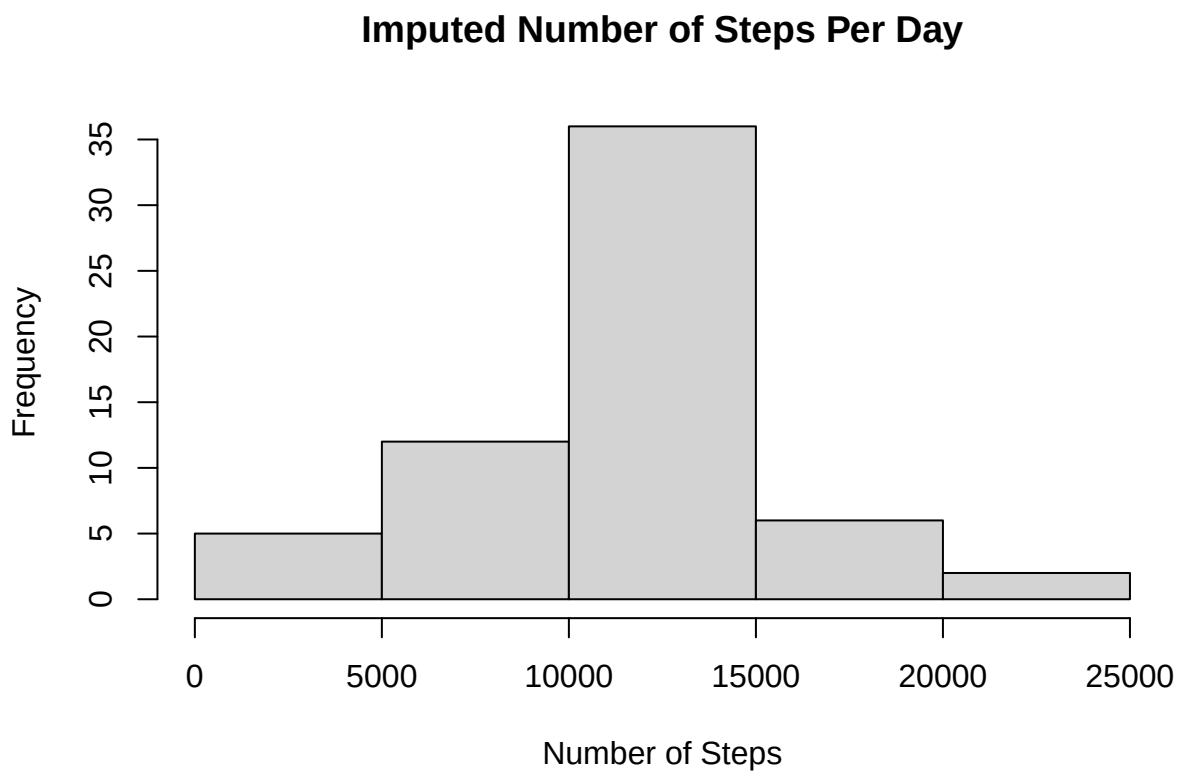
```
missingV <- is.na(Data$steps)
```

Create a new dataset that is equal to the original dataset but with the missing data filled in.

```
miss_Data <- transform(Data, steps = ifelse(is.na(Data$steps),
                                            + meanSteps$steps[match(Data$interval,
                                                                    + meanSteps$interval)],
                                            + Data$steps))
```

Make a histogram of the total number of steps taken each day and Calculate and report the mean and median total number of steps taken per day.

```
iStepsbInt <- aggregate(steps ~ date, miss_Data,FUN=sum)
hist(iStepsbInt$steps, main = "Imputed Number of Steps Per Day", xlab = "Number of Steps")
```



```
impMeanSteps <- mean(iStepsbInt$steps, na.rm = TRUE)
impMedSteps <- median(iStepsbInt$steps, na.rm = TRUE)
diffMean = impMeanSteps - meanSteps
diffTotal = sum(iStepsbInt$steps) - sum(TStepforday$steps)
```

Are there differences in activity patterns between weekdays and weekends?

Create a new factor variable in the dataset with two levels – "weekday" and "weekend"

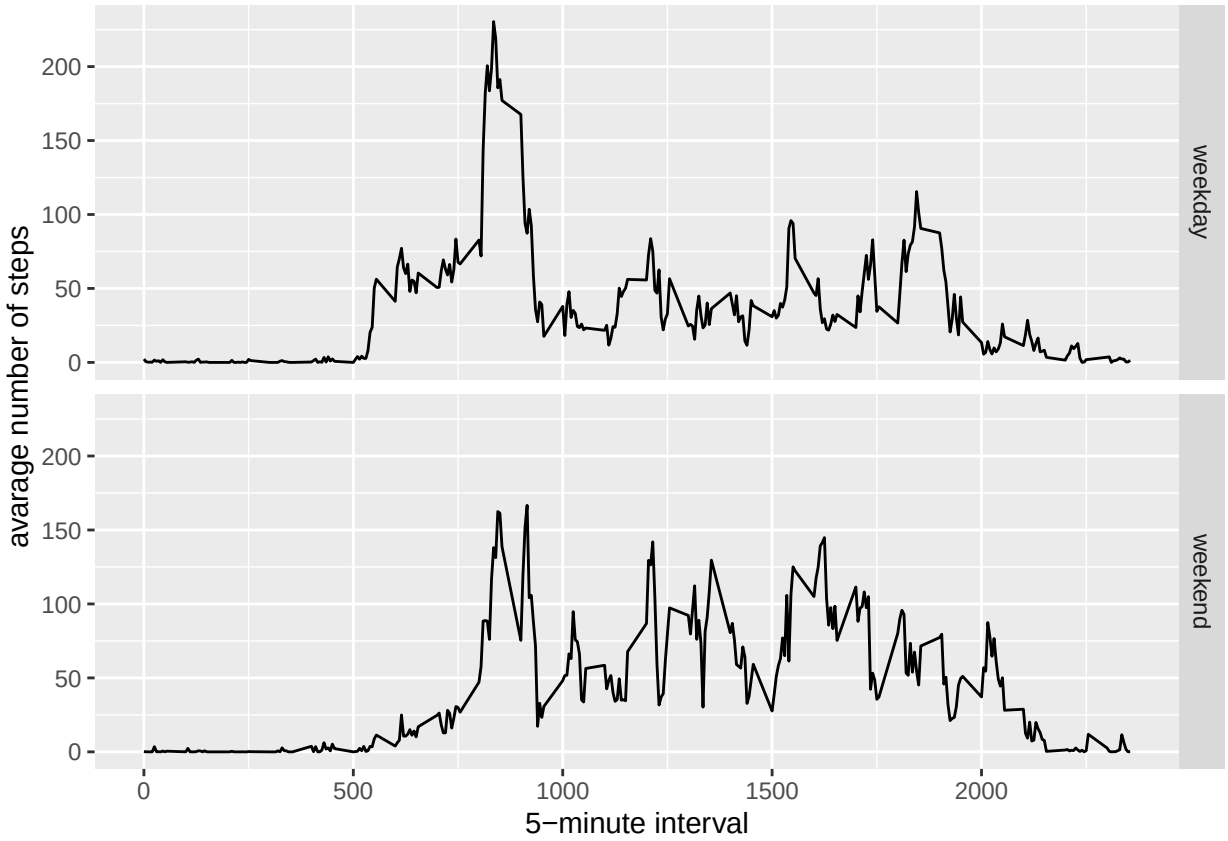
```
DayType <- function(date) {
  day <- weekdays(date)
  if (day %in% c('Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday')) return ("weekeday")
  else if (day %in% c('Saturday', 'Sunday'))
    return ("weekend")
  else stop ("Format no valid") }
```

#For this part the weekdays() function may be of some help here.

```
miss_Data$dateType <-  ifelse(as.POSIXlt(miss_Data$date)$wday %in% c(0,6), 'weekend', 'weekday')
```

Make a panel plot containing a time series plot

```
AVG <- aggregate(steps ~ interval + dateType, data=miss_Data, mean)
ggplot(AVG, aes(interval, steps)) + geom_line() + facet_grid(dateType ~ .) + xlab("5-minute interval") +
```


avarage number of steps
200
150
100
50
0
weekday
weekend
0
500
1000
1500
2000
5−minute interval